

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:41 ; Search time 170.72 Seconds

(without alignments)
23.864 Million cell updates/sec

Title: US-09-331-631a-7_COPY_81_140

Perfect score: 342
Sequence: 1 LQROYQCCGRCQEQQCGOR.....HENYHNKRNSEEEGGOR 60

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR66:**
2: PIR1:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342	100.0	566	2 S22477	vicillin precursor
2	133	38.9	509	2 S08059	alpha-globulin typ
3	133	38.9	605	2 S06398	alpha-globulin typ
4	132	38.6	588	1 FMCNAB	protein PVI00 (imp
5	104	30.4	810	2 T44430	rep protein homolo
6	103.5	30.3	1038	2 T02634	hypothetical prote
7	100	29.2	613	2 S27770	G-box-binding fact
8	93	27.2	708	2 A53185	hypothetical prote
9	93	27.2	930	2 T08588	Mopa box protein -
10	92	26.9	139	2 A26892	protein kinase yak
11	91	26.6	1457	2 T14577	involucrin - dog
12	90.5	26.5	285	1 T46207	sex-determining pr
13	90.5	26.5	395	2 S35565	sex-determining pr
14	90.5	26.5	395	2 S43344	regulatory protein
15	90	26.3	905	1 RGVYSS	protein kinase - s
16	90	26.3	1094	2 S49313	conserved hypotet
17	89	26.0	646	2 D82493	hypothetical prote
18	89	26.0	853	2 T46347	Lotia-like protein
19	89	26.0	1010	2 T13167	sensory transducti
20	88	25.7	2150	2 S71629	prunin 1 precursor
21	87	25.4	551	2 S51941	regulator protein
22	87	25.4	1154	2 S69206	mezoletic antigen
23	86	25.1	255	2 A60637	legumin B - fava b
24	86	25.1	411	2 T29475	unconventional myo
25	86	25.1	564	2 S37241	A-kinase anchor pr
26	85.5	25.0	1737	2 A59235	homeotic protein p
27	85.5	25.0	2359	2 T03094	gene mastermind pr
28	85	24.9	1403	2 S24548	
29	85	24.9	1655	2 T13998	

30	84.5	24.7	648	1 J01150	protein kinase (EC
31	84	24.6	1761	2 T13675	hypothetical prote
32	83.5	24.4	544	1 I36911	involucrin L - dou
33	83.5	24.4	905	2 T23229	hypothetical prote
34	83.5	24.4	919	2 A39248	androgen receptor
35	83	24.3	1090	2 A41696	regulatory protein
36	83	24.3	1180	2 S69205	stripe a/b protein
37	82.5	24.1	467	1 A49377	involucrin - mouse
38	82	24.0	1390	2 T14004	trfa protein - sil
39	82	24.0	1905	2 T18267	multidrug resistan
40	81.5	23.8	4857	2 T03455	ALR protein - huma
41	81.5	23.8	5262	2 T03454	ALR protein - huma
42	81	23.7	339	1 TWH02D	transcription init
43	81	23.7	582	2 B53234	vicillin-like stora
44	81	23.7	758	2 S54522	hypothetical prote
45	81	23.7	816	2 S05548	gap protein hunchb

ALIGNMENTS

RESULT 1

S22477 vicillin precursor - cacao

C:Species: Theobroma cacao (cacao)

C:Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 17-Mar-1999

C:Accession: S22477; S22478; S18105; S22050

R:McHenry, L., Fritz, P.J., Plant Mol. Biol. 18, 1173-1176, 1992

A:Title: Comparison of the structure and nucleotide sequence of vicillin genes of coco

A:Reference number: S22477; MUID:92288309

A:Accession: S22477

A:Molecule type: DNA

A:Residues: 1-566 <MCCH>

A:Cross-references: EMBL:X62625

A:Accession: S22478

A:Molecule type: mRNA

A:Residues: 1-452 <MC2>

A:Cross-references: EMBL:X62626

A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1

C:Superfamily: glycinn

C:Keywords: seed; storage protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-566/Product: vicillin #status predicted <MAT>

Query Match Best Local Similarity 100.0%; Pred. No. 1.3e-25;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 81 LQROYQCCGRCQEQQCGORCQRCWEYKEQERGEHENYHNKRNSEEEGGOR 140

RESULT 2 S08059

alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)

M:Alternate names: Seed storage protein

C:Species: Gossypium hirsutum (upland cotton)

C:Date: 31-Mar-1990 #sequence-revision 31-Mar-1990 #text-change 30-Sep-1993

C:Accession: S08059

R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.

A:Reference number: S06398

A:Accession: S08059

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-509 <CHI>

C:Superfamily: glycinn

```

Query Match 38.9%; Score 133; DB 2; Length 509;
Best Local Similarity 35.8%; Pred. No. 1e-05;
Matches 29; Conservative 12; Mismatches 18; Indels 22; Gaps 3;

Oy 2 0ROYOCCGRCGCEGEOGCEGEOCCGRCGCEGCEGEO-----BRGEHENYHN 46
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 40 0ROFCEGCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCHCH 99
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Oy 47 ----HK---NRSEEGGQR 60
||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 100 PFHFHRSFOSKRFREHGNFR 120
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 3
alpha-globulin type A precursor - upland cotton
506398
N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S063398
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A:Reference number: 506398
A:Accession: S06911
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.
C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-588/Product: alpha-globulin storage proprotein #status predicted <MAT>
F:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.9%; Score 133; DB 2; Length 605;
Best Local Similarity 43.3%; Pred. No. 1.2e-05;
Matches 26; Conservative 14; Mismatches 18; Indels 2; Gaps 2;

Oy 3 ROYOCCGRCGCEGEOGCEGEOCCGRCGCEGCEGEO--RGEHEN-YHNHKNRSEEGGQR 60
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 120 KQFKCQGRQCMQGRPRPRKQCVKCEHQYQEDPWKGRKNKREBEESDEGEGQR 179
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 4
FWCNAB
alpha-globulin B precursor (clone C72) - upland cotton
N:Alternate names: seed storage protein; vicilin precursor
C:Species: Gossypium hirsutum (upland cotton)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: A30838; S06911
R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII CE
A:Reference number: A30838
A:Accession: A30838
A:Molecule type: mRNA
A:Residues: 1-588 <GBL>
A:Cross-references: GB:M16891; NID:g167374; PIDN:AAA33071.1; PID:g167375
A:Experimental source: var. Coker 201
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A:Reference number: 506398
A:Accession: S06911
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.
C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-588/Product: alpha-globulin storage proprotein #status predicted <MAT>
F:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match Similarity      38.6%; Score 132; DB 1; Length 588;
Best Local Similarity      35.8%; Pred. No. 1.4e-05;
Matches      29; Conservative      11; Mismatches      19; Indels      22; Gaps      3;

QY      2   ORGYOCCGRCOEEOOGCQRKCKMEQYKED-----ERGENENYN 46
          |||::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db       121  QRFQCCEHCIOQCQRREBRKCQCVRCCRERYCNPMREREAEAEETDEGEQESHN 180
QY      47   ----HK---NRSEEEGOOR 60
          |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db       181  PFHFRRSFQSFRREEHGNFR 201

RESULT      5
T44430
Protein PV100 [imported] - winter squash
C:Species: Cucurbita maxima (winter squash)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44430
R.Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A>Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin
A:Reference number: Z22767; MUID:99107919
A:Accession: T44430
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-810 <YAM>
A:Cross-references: EMBL:AB019195; NID:g3808061; PID:NBA43056.1; PID:g3808062

Query Match      30.4%; Score 104; DB 2; Length 810;
Best Local Similarity      31.6%; Pred. No. 0.0088;
Matches      18; Conservative      18; Mismatches      21; Indels      0; Gaps      0;

QY      4   OYOOCOCGRCOEQOCQRKCKMEQYKEERGEHENYHNKKNRSEEEGOOR 60
          |:|::||| |::| |::||| |::||| |::||| |::||| |::|||
Db       77  EYEVCRLKQVAERKEVEOQRKCEYCERLREREGGEDVDEVRRDPPEWEREBOR 133

RESULT      6
T02634
rep protein homolog - slime mold (Dictyostellium discoideum) plasmid dtps
C:Species: Dictyostellium discoideum
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 29-Oct-1999
C:Accession: T02634
R.Ribben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.D.; Kiyosawa, H.; H
A>Title: Dictyostellium discoideum nuclear plasmid dtps is a chimera related to the dd
A:Reference number: Z14684; MUID:99198836
A:Accession: T02634
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1038 <RIE>
A:Cross-references: EMBL:AF000580; NID:g3068582; PID:NAA14390.1; PID:g3068583
A:Experimental source: strain WS2162
C:Genetics:
A:Gene: rep
A:Genome: plasmid
A:Mobile element: plasmid dtps

Query Match      30.3%; Score 103.5; DB 2; Length 1038;
Best Local Similarity      37.3%; Pred. No. 0.012;
Matches      22; Conservative      17; Mismatches      19; Indels      1; Gaps      1;

QY      2   ORGYOCCGRCOEEOOGCQRKCKMEQYKEDRGEGHENYHNKKNRSEEEGOOR 60
          |::||| |::||| |::||| |::||| |::||| |::||| |::|||
Db       940  OEQGCRERRREQEQEQEQEQEQ--QEQQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 997

RESULT      7

```


C:Species: *Canis lupus familiaris* (dog)
C:Date: 16-Aug-1996 #sequence_revision 23-Aug-1996 #text_change 22-Jun-1999
C:Accession: I46207
R:Tseng, H.; Green, H.
Mol. Biol. Evol. 7, 293-302, 1990
A:Title: The involucrin genes of pig and dog: comparison of their segments of repeats with
A:Reference number: I46207; MUID:90348475
A:Accession: I46207
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-285 <TSK>
C:Cross-references: GB:M34442; NID:G163980; PIDN:AAA30853.1; PID:G163981
C:Comment: During the terminal differentiation of keratinocytes, this protein from the c
linked envelope under the plasma membrane.
C:Superfamily: involucrin
C:Keywords: cornified cell envelope; duplication; epidermis; tandem repeat

Query Match	26.58;	Score 90.5;	DB 1;	length 285;
Best Local Similarity	31.7%;	Pred. No. 0.069;		
Matches 19;	Conservative 18;	Mismatches 18;	Indels 5;	Gaps 2

```

Qy      6  QCCGRCGCEGGGGRGQCCQCC---RKCMQYKQDGEHFNHNNH--KRNSEEEGGQR 60
      ||| : |||| | : | : : ||| : | : : : | : | :
Db      73  QCCCEGCEGGGQKQGESQCKLHLCGLEQHNGEGESGQCKLYPCQLCEGGQEGEGEGDQ 132

```

RESULT 13

sex-determining protein SRY - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence-revision 19-Mar-1997 #text-change 31-Jan-2000
C:Accession: S35565; S10938; A59162
R:Tucker, P.K.; Lundrigan, B.L.
Nature 364, 715-717, 1993
A>Title: Rapid evolution of the sex determining locus in Old World mice and rats
A:Reference number: S35565; MUID:93361118

A: Accession: S35565
A: Molecule type: DNA
A: Residues: 1-395 <TUC>
A: Experimental source: liver
R: Gubbay, J.: Collignon, J.: Koopman, P.: Capel, B.: Economou, A.: Muensterberg, A.: Vivian, J.: Bishop, C.A.: Melnick, M.L.: Tjian, R.: Gurley, L.: Zannis-Hadjopoulos, D.: Weiss, M.M.: 1987, Nature 326, 245-250, 1990
A: Title: A gene mapping to the sex-determining region of the mouse Y chromosome is a member of the HMG domain superfamily.
Reference number: S10938; MUID:90326154

A;Accession: S10938
A;Molecule type: DNA
A;Residues: 'TGDNCLCS', 1-91 <MOU>
A;Cross-references: GB:X55491; NID:g287804

A;Status: not compared with conceptual translation

A;Residues: 1-125 <GUB>

A; Note: the sequence is revised in GenBank entry MMSRYA, release 113.0

A;Gene: SRY

C; Superfamily: unassigned HMG box proteins; HMG box homology

F;2-77/Domain: HMG box homology <HMG1>

Query Match	26.58; Score 90.5; DB 2; Length 395;
-------------	--------------------------------------

Matches 18; Conservative 20; Mismatches 20; Indels 5; Gaps 1;

02 3 R0Y0Q0GRC0E0Q0G0RE0Q0C-----0RK0WE0YKE0ERGEHENYHNHKNRSEEEEG 57

Db 160 QQQQQFYDHNQQQQQQQQQFNDHNQKQFNDHNQQQQQFNDHNNNNEQFNDHNQ 219

QY 58 QQR 60
11:

Db 220 000 222

RESULT 1.4

sex-determining protein Sry - mouse
C:Species: Mus musculus domesticus (western European house mouse)
C:Date: 25-Dec-1994 #sequence-revision 03-Nov-1995 #text-change 24-Sep-1999
C:Accession: S43344; 578062
R:Coward, P.; Nagai, K.; Chen, D.; Thomas, H.D.; Nagamine, C.M.; Lau, Y.F.C.
Nature Genet. 6, 245-250, 1994
A:Title: Polymorphism of a CAG trinucleotide repeat within Sry correlates with B6.YJID
#:Reference number: S43344; M0JID:94282071

A: Accession: S43344
A: Molecule type: DNA
A: Residues: 1-395 <COM>
A: Experimental source: liver: strain Torino
R: Coward, P.; Nagai, K.; Chen, D.; Thomas, H.D.; Nagamine, C.M.; Lau, Y.F.C.
submitted to the EMBL Data Library, November 1993
A: Description: Polymorphism of a CAG trinucleotide repeat within SRY correlates with
A: Reference number: S78062

A:Accession: 578062
A:Molecule type: DNA
A:Residues: 1-62, 'T', 64-132, 'L', 134-142, 'P', 146-168, 171-208, 'Q', 210, 'Q', 212-234 <CONF>
A:Cross-references: EMBL:U03645; NID:9460122; PIDD:AA060446.1; PID:9460123

```
A: Experimental source: liver; strain Torino
A: Genetics:
A: Gene: sly
A: Map position: y
C: Superfamily: unassigned HMG box proteins; HMG box homology
C: Keywords: DNA binding
F: 2-77/Domain: HMG box homology<HMG1>
F: 144-366/Region: glutamine-rich
```

Query Match	26.58;	Score 90.5;	DB 2;	Length 395;
Best Local Similarity	28.68;	Pred. No. 0.091;		
Matches 18;	Conservative 20;	Mismatches 20;	Indels 5;	Gaps 1;

```
0y      3 R0T0R0C0GRC0E0R0G0R0E0R0C-----0RKSC0E0YK0ED0RG0EN0Y0NNHKNRSEEEEG 57
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     160 QQQQQRFYDNNHQQQQQQQQQQQQQQFNDNHNQKKQGFNDNHNQOQQQGFNNNNNNNNNQEQFNDDNNQ 219
```

Qy	58	QQR	60
		:	
Db	220	QQQ	222

RESULT 15

regulatory protein SNF5 - yeast (*Saccharomyces cerevisiae*)

C:Species: *Saccharomyces cerevisiae*

C;Accession: S44551; S46171; A36375; S12067; S39145

Yeast 10(Suppl.A), S47-S62, 1994

A;Reference number: S44537; MUID:94378722

A;Status: translation not shown

A;Residues: 1-905 <HOL>

R;Brandt, T.; Christiansen, C.; Holmstroem, K.; Kallesoe, T.

A;Reference number: S46157

A;molecule type: DNA

A;Cross-references: EMB

Mol. Cell. Biol. 10, 5616-5625, 1990

